Attorney Docket No.: 3655/0138PUS1 PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant(s): SADOT, Emek	Conf. No.: 2558
Application No.: 10/072,364	Art Unit: 2194
Filed: February 6, 2002	Examiner: ANYA, Charles
Title: CLIENT-CONTROLLED LOAD BALANCER	

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

APPLICANT'S APPEAL BRIEF

Sir:

A Notice of Appeal was filed in the above application on January 27, 2009. This Brief is being filed within two months of the date of the Notice of Appeal as required by 37 C.F.R. 41.37 together with the required fee.

I. REAL PARTY IN INTEREST

The real party in interest in the above-captioned application is Avaya Communication Israel Ltd. as shown by the assignment recorded at patent Reel 012586, Frame 0847 on February 6, 2002.

II. RELATED APPEALS AND INTERFERENCES

There are no prior or pending appeals, interferences or judicial proceedings known to appellant, the appellant's legal representatives or assignee which may be

related to, directly affect or be directly affected by or have a bearing on the Board's decision in the pending appeal.

III. STATUS OF CLAIMS

Claims 1-54 are pending in the subject application. Claims 1-54 are rejected. The rejections of claims 1-54 are being appealed.

IV. STATUS OF AMENDMENTS

No amendments were filed subsequent to the final rejection.

V. SUMMARY OF CLAIMED SUBJECT MATTER

Claim 1

Claim 1 recites a method of selecting a server (108 in Figure 1) to represent a virtual server hosted by a plurality of servers. The method includes providing by a load balancer (102 in Figure 1) not associated with the virtual server values for one or more parameters of two or more paths, each path defined between a point in a vicinity of a client accessing the virtual server and one of the plurality of servers representing the virtual server (page 3, lines 1-10). The method also includes selecting a server to provide data for the client responsive to the values of the one or more parameters (page 3, lines 10-12). The load balancer (102 in Figure 1) comprises a client-controlled load balancer that directly selects the one of the plurality of servers representing the virtual server based on the one or more parameters (page 3, lines 10-12; page 10, line 24 - page 11, line 6).

Claim 24

Claim 24 recite a method of selecting a server to be accessed that includes receiving by a load balancer(102 in Figure 1) a message relating to a virtual server, hosted by a plurality of servers (page 3, lines 1-2), and to a client desiring to receive data from the virtual server (page 3, lines 19-24). The method also includes selecting by the load balancer one of the plurality of servers to provide data to the client based on one or more parameters related to a path to the client (page 3, lines 10-12). The load balancer is closer to the client than to the selected server (page 6, lines 5-10; page 8, lines 24-28), and the load balancer comprises a client-controlled load balancer that directly selects the one of the plurality of servers representing the virtual server based on the one or more parameters (page 3, lines 1-12; page 10, line 24 - page 11, line 6).

Claim 37

Claim 37 recites a method of selecting a server to be accessed that includes receiving by a load balancer (102 in Figure 1) a message relating to a virtual server hosted by a plurality of servers (page 3, lines 1-2) and to a client desiring to receive data from the virtual server (page 3, lines 19-24). The method also includes selecting by the load balancer one of the plurality of servers to provide data to the client (page 3, lines 10-12) at least partially responsive to the cost of communications between the client and one or more of the plurality of servers (page 10, lines 30-33). The load balancer comprises a client-controlled load balancer that directly selects said one of the plurality of servers representing the virtual server based on said one or more

parameters (page 3, lines 1-12; page 10, line 24 - page 11, line 6).

Claim 41

Claim 41 recites a load balancer (102 in Figure 1) that includes an interface adapted to receive server access messages from clients (page 9, lines 3-7 and Figure 1) and a processor adapted to determine for at least one of the messages whether the message requires load balancing responsive to at least one attribute different from the identity of the server referenced by the message (page 10, lines 1-17). The processor is adapted to select for at least one message determined to require load balancing a server to service the client (page 7, lines 4-9; page 9, lines 3-7; Figure 4). The processor comprises a client-controlled processor that directly selects the server to service the client based on the at least one attribute (page 3, lines 1-12; page 10, line 24 - page 11, line 6).

Claim 47

Claim 47 recites a method of selecting a server to be accessed that includes receiving by a load balancer (102 in Figure 1) a message relating to a virtual server hosted by a plurality of servers (page 3, lines 1-2) and to a client desiring to receive data from the virtual server (page 3, lines 19-24). The method also includes choosing a function from a plurality of predetermined functions utilized by the load balancer for selecting servers, responsive to the received message (page 7, lines 16-20). Also, selecting, by the load balancer, one of the plurality of servers that minimizes or maximizes the chosen function, to provide data to the client (page 7, lines 21-22). The

load balancer comprises a client-controlled load balancer that directly selects said one of the plurality of servers representing the virtual server that minimizes or maximizes the chosen function page 3, lines 1-12; page 10, line 24 - page 11, line 6).

Claim 52

Claim 52 recites a method of selecting a server to be accessed by a client via a wide area network (WAN) (110 in Figure 1) from among a plurality of servers associated with a domain name (page 8, lines 12-24). The method includes providing a client-controlled load balancer (102 in Figure 1) in a local area network (LAN) (104 in Figure 1) connected to the WAN, the LAN including the client. Also receiving at the load balancer a list of addresses of servers hosting the domain name (page 8, lines 12-23) and selecting by the load balancer one of the addresses of the plurality of servers based on a parameter related to a path between a point in the vicinity of the client and one of the plurality of servers (page 10, lines 24-33).

VI. GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL

Whether claims 1-4, 6, 7, 13-17, 24-33, 35-37, 41, 42 and 44[1] are properly rejected under 35 U.S.C. 103(a) as being unpatentable over U.S. 7,047,301 (hereinafter "Skene") in view of U.S. Patent No. 6,182,139 (hereinafter "Brendel").

Whether claims 8-10, 18-23, 34, 38-40, 43 and 47-51 are properly rejected under 35 U.S.C. 103(a) as being unpatentable over Skene in view of Brendel and further in

[1] Claim 5 is not mentioned in any of the claim rejections but is discussed in connection with the claims that are rejected based on Skene in view of Brendel. It is therefore believed that the examiner is rejecting claim 5 as being unpatentable over Skene in view of Brendel. Whether claim 5 is properly rejected as being unpatentable over Skene in view of Brendel is also a ground of rejection to be reviewed on appeal.

view of U.S. Patent No. 6,249,801 (hereinafter "Zisapel").

Whether claims 11, 12, 45 and 46 are properly rejected under 35 U.S.C. 103(a) as being unpatentable over U.S. 2001/0047415 (the publication of the Skene patent referred to above and also referred to herein as "Skene") in view of Brendel and further in view of U.S. 6,389,462 (hereinafter "Cohen").

Whether claims 52 and 53 are properly rejected under 35 U.S.C. 103(a) as being unpatentable over Brendel in view of U.S. Patent No. 6,185,601 (hereinafter "Wolff").

Whether claim 54 is properly rejected under 35 U.S.C. 103(a) as being unpatentable over Brendel in view of Wolff and further in view of U.S. 2001/0039588 (hereinafter, "Primak").

VII. ARGUMENT

A. REJECTIONS BASED ON SKENE IN VIEW OF BRENDEL

Independent Claim 1

Claim 1 is rejected under 35 U.S.C. 103(a) as being unpatentable over Skene in view of Brendel. Claim 1 is directed to a method of selecting a server to represent a virtual server hosted by a plurality of servers. The method involves a load balancer that is not associated with the virtual server. The load balancer provides values for one or more parameters of two or more paths between a point in a vicinity of a client and one of the plurality of servers representing the virtual server. One of the servers is selected to provide data to the client based on the values of the parameters. The load balancer is a client-controlled load balancer that directly selects the one of the plurality of servers representing the virtual server based on the one or more parameters.

Skene is directed to a more-or-less conventional load balancing system. Skene includes virtual servers 110, 111, 112 controlled by server array controllers (SAC) 101, 102, 103. An extended domain name server (EDNS) 160 provides load balancing among the SAC's and helps ensure that requests from a given client get routed consistently to a particular server so that client-specific information stored on a particular server persists over time. Brendel on the other hand, teaches a "client-side dispatcher" that sits on a client machine and intercepts communications from an application running on the machine before they reach the internet. When a given web site is hosted by a plurality of servers, the client side dispatcher helps direct requests from the client to a particular server. Significantly, Brendel acknowledges that his client-side dispatcher cannot perform traditional load balancing because it does not have information about server loads or requests from other clients (column 5, lines 30-35). However, Brendel's dispatcher can perform one very limited load balancing function: the dispatcher can send requests to multiple servers and select the first server that responds as the best server to use (column 5, lines 36-51).

The rejection of claim 1 indicates that the system of Skene should be modified with the teaching of Brendel but provides no description of what modification to Skene is being proposed. The rejection therefore fails to satisfy the requirements of MPEP 706.02(j) which requires the examiner to identify the proposed modification of the applied reference(s) necessary to arrive at the claimed subject matter. The rejection therefore does not present a prima facie case of obviousness. The response to arguments section of the final Office Action, however, seems to indicate that the examiner is proposing to replace one of Skene's clients 150 with Brendel's client side

dispatcher. As discussed below, even if this were done, the result would not be the invention of claim 1.

Claim 1 indicates that a load balancer not associated with a virtual server provides values for at least one parameter of two or more paths. Placing Brendel's client-controlled dispatcher into the system of Skene, however, does not change the fact that Brendel can merely determine which server is the quickest to respond to a request. Brendel does not determine any parameters for any path; his system merely selects the server that responds first. Brendel does not even obtain parameters for the single path of the server that responds first; there is no indication, for example, that a response time is calculated. In addition, no parameters are provided for the paths that are not elected. Skene's EDNS may perform traditional load balancing, but it is not a client-controlled load balancer. Brendel's client controlled dispatcher provides no path parameters. Placing Brendel's client-controlled dispatcher into Skene's system therefore does not result in or render obvious the invention of claim 1.

Claims 2, 3, 6, 7, 13-17

Claims 2, 3, 6, 7, and 13-17 depend from claim 1 and are submitted to be allowable for at least the same reasons as claim 1.

Claim 4

Claim 4 further provides that the at least one parameter of the two or more paths includes one of jitter, round trip delay or a hop count. The examiner refers to Skene to show various parameters that are used by traditional server side load balancers.

However, this in no way shows or suggests that these parameters can be provided by Brendel's client side dispatcher if it is added to the Skene system. The fact that a server side load balancer can determine various parameters in no manner suggests that a client side dispatcher (which has no information regarding server load, etc.) can perform these functions. Adding Brendel's dispatcher to Skene's system does not render obvious the method of claim 4, and claim 4 is submitted to further distinguish over Skene and Brendel for this reason.

Claim 5

Claim 5 further provides that the one or more parameters of the two or more paths comprise a cost of communication. Skene indicates that a traditional server side load balancer can rely on "other metrics" to perform load balancing functions. The examiner appears to argue that a "cost of communication" is inherent in a discussion of "other metrics." However, a parameter comprising cost of communication is not mentioned in either Skene or Brendel, and the record does not contain arguments to support reliance on the theory of inherency. Claim 5 is submitted to further distinguish over Skene and Brendel for this reason.

Independent Claim 24

Claim 24 recites a method of selecting a server to be accessed that involves a client-controlled load balancer. The method includes selecting, by the load balancer, one of a plurality of servers to provide data to the client based on one or more parameters related to a path to the client. If Brendel's client side dispatcher is placed

into Skene's system, it will at most be able to determine which one of a plurality of servers responds first to its multiple requests. Nothing in Brendel indicates that Brendel's dispatcher can determine any parameter of any path to a client. Brendel identifies a particular server, the one which responds first, but does not determine information regarding any path and cannot make a selection based on any parameters related to a path (such as round trip delay, jitter, etc.). Claim 24 is submitted to be allowable over Brendel and Skene for at least this reason.

Claims 25-33, 35 and 36

Claims 25-33, 35 and 36 depend from claim 24 and are submitted to be allowable for at least the same reasons as claim 24.

Independent Claim 37

Claim 37 recites a method of selecting a server to be accessed using a client-controlled load balancer. The selecting is at least partially responsive to the cost of communications between the client and one or more of the plurality of servers. Neither Skene nor Brendel mentions a selection based on communications cost, in connection with a traditional load balancer or otherwise. However, the examiner again refers to language in Skene regarding unspecified "metrics" to show that Skene discloses a selection based at least partially on cost of communications.

A selection based on cost of communications is not shown by the references. A selection based on cost of communications is not necessarily present in the references and therefore cannot be inherent in the references (MPEP 2112). At least this limitation

of claim 37 is not satisfied by Skene and Brendel, and claim 37 is submitted to be allowable for at least this reason.

Independent Claim 41

Claim 41 recites, inter alia, a client-controlled load balancer that includes an interface adapted to receive server access messages from clients and a processor. The processor is adapted to determine, for at least one of the messages, whether the message requires load balancing. The determination is based on at least one attribute different from the identity of the server referenced by the message. Without limitation, the attribute may be the identity of the client requesting access to a server (page 10, lines 1-17). This limitation regarding an attribute different from the identity of the server referenced by the message is not addressed in the rejection of claim 41 and is not shown or suggested by Skene or Brendel. Claim 41 is submitted to be allowable over Skene and Brendel for at least this reason.

Claim 42

Claim 42 further defines the at least one attribute recited in claim 41 as the time at which the message is received at the interface. Skene and Brendel do not teach a determination of whether a message requires load balancing based on an attribute different from the identity of a server. Skene and Brendel therefore also do not teach a determination based on the time at which the message is received at an interface as an attribute on which a determination is based. Claim 42 further distinguishes over Skene and Brendel for this reason.

Claim 44

Claim 44 further defines the at least one attribute recited in claim 41 as a protocol to govern the communication with the server. Skene and Brendel do not teach a determination of whether a message requires load balancing based on an attribute different from the identity of a server. Skene and Brendel therefore also do not teach a determination based on a protocol to govern the communication with the server as an attribute on which a determination is based. Claim 44 further distinguishes over Skene and Brendel for this reason.

B. REJECTIONS BASED ON SKENE IN VIEW OF BRENDEL AND FURTHER IN VIEW OF ZISAPEL

Claims 8-10

Claim 1 includes the limitation: providing, by a load balancer not associated with the virtual server, values, for one or more parameters, of two or more paths, each path defined between a point in a vicinity of a client accessing the virtual server and one of the plurality of servers representing the virtual server. Claim 8 further recites that providing values for one or more parameters comprises measuring at least one of the parameters.

As an initial matter, it is submitted that a prima facie case of obviousness has not been presented in connection with claim 8. Section 706.02(j) of the MPEP provides that in order to reject a claim under 35 U.S.C. 103(a), the examiner should set forth "the proposed modification of the applied reference(s) necessary to arrive at the claimed subject matter" The Office Action only indicates that it would have been "obvious to

modify the system of Brendel and Skene with the teaching of Zisapel…." Stating that some modification should be made without explaining what modification is being proposed does not satisfy this requirement. Claim 8 is submitted to be allowable for at least this reason.

Furthermore, to the extent the references can be combined, they do not show or suggest the invention of claim 8. As discussed in connection with claim 1, Skene and Brendel do not show or suggest providing parameters of two or more paths, and Zisapel does not address this shortcoming of Skene and Brendel. Zisapel indicates that certain parameters may be measured in traditional load balancers. However, nothing in the record shows or suggests how this teaching would affect Brendel's client side dispatcher, which is not capable of performing traditional load balancing. Claim 8 is submitted to be allowable for this reason as well.

Claims 9 and 10 depend from claim 8 and are submitted to be allowable for at least the same reasons as claim 8.

Claims 18-23

Claim 18 depends from claim 1 and indicates that selecting a server comprises choosing a function of the one or more parameters to be minimized and selecting a server which minimizes the chosen function. The examiner acknowledges that Skene and Brendel do not satisfy this limitation. Zisapel discusses functions of a traditional load balancer. Nothing in the record shows or suggests that such functions could be performed by Brendel's client side dispatcher. Claim 18 further distinguishes over the art of record for this reason.

Claims 19-23 depend from claim 18 and are submitted to be allowable for at least the same reasons as claim 18.

Claim 34

Claim 34 depends from claim 24 and recites that selecting one of the servers comprises selecting a server which has a lowest cost path to the load balancer. Neither Skene nor Brendel discusses path costs. Zisapel's reference to a "total weighted value" does not comprise a determination of path cost. Zisapel does not address the shortcomings of Skene and Brendel, and claim 34 is submitted to be allowable for at least this reason.

Claims 38-40

Claims 38-40 depend from claim 37. Claim 37 recites, inter alia, selecting, by the load balancer, one of the plurality of servers to provide data to the client, at least partially responsive to the cost of communications between the client and one or more of the plurality of servers. Each of claims 38-40 further defines the act of selecting as being at least partially responsive to a cost of communications. The phrase "lowest weighted total" referred to by the examiner does not show or suggest a determination based on communications costs, and Zisapel does not address the shortcomings of Skene and Brendel. Claims 38-40 are submitted to be allowable over the art of record for at least this reason.

Claim 43

Claim 41 recites load balancing responsive to at least one attribute different from the identity of the server referenced by the message. Claim 43 indicates that the attribute comprises the identity of the client. Zisapel mentions at column 5, line 32-38 that the source address of a request may be known. However, Zisapel in no manner suggests taking any action based on this source address. Zisapel does not address the shortcomings of Skene and Brendel, and claim 43 is submitted to be allowable for at least this reason.

Independent claim 47

Claim 47 recites, inter alia, a method of selecting a server to be accessed that includes receiving by a load balancer a message relating to a virtual server hosted by a plurality of servers, and to a client desiring to receive data from the virtual server. Claim 47 further recites choosing a function, from a plurality of predetermined functions utilized by the load balancer for selecting servers, responsive to the received message, and selecting by the load balancer one of the plurality of servers that minimizes or maximizes the chosen function. Skene indicates that traditional load balancers operate according to a variety of load balancing metrics. Neither Skene nor the other references suggest in any manner that the metric selected is "responsive to the received message" as recited in claim 47. Claim 47 is submitted to be allowable over the art of record for this reason.

Claim 48

Claim 48 further defines how a function is chosen, namely that it is chosen

responsive to an identity of the client. The art of record provides no information regarding choosing a function responsive to a received message or choosing a function responsive to an identity of a client. Claim 48 is submitted to be allowable over the art of record for at least this reason.

Claim 49

Claim 49 further defines how a function is chosen, namely that it is chosen responsive to a time at which the message is received. The art of record provides no information regarding choosing a function responsive to a received message or choosing a function responsive to a time at which a message is received. Claim 49 is submitted to be allowable over the art of record for at least this reason.

Claims 50 and 51

Claims 50 and 51 depend from claim 47 and are submitted to be allowable for at least the same reasons as claim 47.

C. REJECTIONS BASED ON SKENE AND BRENDEL AND FURTHER IN VIEW OF COHEN

Claim 11

Claim 11 depends from claim 1 and recites an act of changing the destination IP address of packets received by a load balancer from a client to an IP address of a selected server. The Office Action acknowledges that this is not shown by Skene and Brendel, which references were used as the basis for rejecting claim 1. The examiner

cites Cohen to cure this deficiency. Cohen teaches a method of directing web requests to proxy caches. The examiner indicates that the reason for modifying Skene and Brendel based on Cohen is to allow "a client to transparently establish a TCP connection with proxy cache." Skene and Brendel do not discuss proxy caches. It is therefore not clear why Skene or Brendel should be modified to establish connections with proxy caches or how such a modification would result in the invention of claim 11. A prima facie case of obviousness has not been provided in connection with claim 11, and claim 11 is submitted to be allowable for at least this reason.

Claim 12

Claim 12 depends from claim 1 and further recites changing the source IP address of packets received by the load balancer from the selected server. The examiner indicates that this is not shown by Skene or Brendel but argues that Cohen's proxy caching method makes a change to Skene and Brendel obvious. Skene and Brendel are not directed to proxy caching systems, and adding proxy caching functionality to Skene and Brendel would not result in the invention of claim 12. Claim 12 is submitted to be allowable for at least this reason.

Claims 45 and 46

Claims 45 and 46 depend from claim 41. Cohen does not address the shortcomings of Skene and Brendel discussed above in connection with claim 41. Claims 45 and 46 are therefore submitted to be allowable over the art of record for at least this reason. Moreover, claims 45 and 46 further distinguish over the art of record

for the reasons provided above in connection with claims 11 and 12, namely, that nothing in Cohen's proxy caching method suggests any modification to Skene and Brendel that would result in a load balancer as claimed.

Independent claim 52

Claim 52 is rejected under 35 U.S.C. 103(a) as being unpatentable over Brendel in view of Wolff. Claim 52 recites a method of selecting a server to be accessed by a client via a wide area network (WAN) from among a plurality of servers associated with a domain name. The method includes providing a client-controlled load balancer in a local area network (LAN) connected to the WAN, the LAN including the client, and receiving at the load balancer a list of addresses of servers hosting the domain name. The method also includes selecting by the load balancer one of the addresses of the plurality of servers based on a parameter related to a path between a point in the vicinity of the client and one of the plurality of servers. Brendel does not make determinations based on parameters related to paths. Brendel merely selects the server that responds first to a request. Wolff does not address this shortcoming of Brendel, and claim 52 is submitted to be allowable for at least this reason.

Claim 53

Claim 53 depends from claim 52 and is submitted to be allowable for at least the same reasons as claim 52.

Claim 54

Claim 54 is rejected under 35 U.S.C. 103(a) as being unpatentable over Brendel in view of Wolff and further in view of Primak. Claim 54 depends from claim 52. Primak does not address the shortcomings of Brendel and Wolff discussed above in connection with claim 52. Claim 54 is therefore submitted to be allowable for at least the same reasons as claim 52.

CONCLUSION

Reconsideration and allowance of claims 1-54 is earnestly solicited in view of the foregoing arguments.

To the extent necessary, a petition for an extension of time under 37 C.F.R. 1.136 is hereby made. Please charge any shortage in fees due in connection with the filing of this, concurrent and future replies, including extension of time fees, to Deposit Account 50-3828 and please credit any excess fees to such deposit account.

Respectfully submitted,

/Scott T Wakeman #37750/

Scott L. Lowe
Registration No. 41,458
Scott T. Wakeman
Registration No. 37,750

PO BOX 1364
Fairfax,VA 22038-1364
1.703.621.7140

Date: March 25, 2009

VIII. Claims Appendix

1. A method of selecting a server to represent a virtual server hosted by a plurality of servers, comprising:

providing, by a load balancer not associated with the virtual server, values, for one or more parameters, of two or more paths, each path defined between a point in a vicinity of a client accessing the virtual server and one of the plurality of servers representing the virtual server; and

selecting a server to provide data for the client, responsive to the values of the one or more parameters,

wherein the load balancer comprises a client-controlled load balancer that directly selects said one of the plurality of servers representing the virtual server based on said one or more parameters.

2. A method according to claim 1, wherein the load balancer and the client are in the same metropolitan area.

3. A method according to claim 1, wherein the client-controlled load balancer resides between the client and the virtual server.

4. A method according to claim 1, wherein the one or more parameters comprise at least one of a jitter, a round trip delay or a hop count.

5. A method according to claim 1, wherein the one or more parameters comprise a cost of communication.

6. A method according to claim 1, wherein selecting the server comprises selecting, by the client-controlled load balancer, responsive to receiving identification of a virtual server requested by the client.

7. A method according to claim 6, wherein selecting the server comprises selecting, by the client-controlled load balancer, responsive to receiving a connection establishment request from the client.

8. A method according to claim 6, wherein providing the values for the one or more parameters comprises measuring at least one of the parameters.

9. A method according to claim 8, wherein measuring at least one of the parameters, for at least one of the paths, is performed before receiving the connection establishment request.

10. A method according to claim 8, wherein measuring at least one of the parameters for at least one of the paths is performed after receiving the connection establishment request.

11. A method according to claim 1, further comprising changing the destination IP address of packets received by the load balancer from the client, to an IP address of the selected server.

12. A method according to claim 1, further comprising changing the source IP address of packets received by the load balancer from the selected server.

13. A method according to claim 1, further comprising transmitting an IP address of the selected server to the client.

14. A method according to claim 13, wherein transmitting the IP address of the selected server to the client comprises transmitting a DNS response.

15. A method according to claim 1, wherein ones of the plurality of servers are located in different geographical regions.

16. A method according to claim 1, wherein selecting a server to provide data for the client comprises selecting, by the load balancer, a second load balancer which is to perform the server selection and selecting, by the second load balancer, a server to provide data for the client.

17. A method according to claim 1, wherein the virtual server hosts a web site.

18. A method according to claim 1, wherein selecting a server to provide data for the client comprises selecting a server which minimizes a function of the one or more parameters.

19. A method according to claim 18, wherein selecting a server to provide data comprises choosing a function of the one or more parameters to be minimized and selecting a server which minimizes the chosen function.

20. A method according to claim 19, wherein the function is chosen responsive to a protocol with which the virtual server is accessed.

21. A method according to claim 19, wherein the function is chosen responsive to the virtual server accessed.

22. A method according to claim 19, wherein the function is chosen responsive to an attribute of the client.

23. A method according to claim 19, wherein the function is chosen responsive to the time of the selection.

24. A method of selecting a server to be accessed, comprising:

receiving, by a load balancer, a message relating to a virtual server, hosted by a plurality of servers, and to a client desiring to receive data from the virtual server; and

selecting, by the load balancer, one of the plurality of servers to provide data to the client based on one or more parameters related to a path to the client,

wherein the load balancer is closer to the client than to the selected server, and

wherein the load balancer comprises a client-controlled load balancer that directly selects said one of the plurality of servers representing the virtual server based on said one or more parameters.

25. A method according to claim 24, wherein the load balancer is closer to the client than to any of the plurality of servers hosting the virtual server.

26. A method according to claim 24, wherein the load balancer is in the same metropolitan area as the client.

27. A method according to claim 24, wherein the client-controlled load balancer resides between the client and the virtual server.

28. A method according to claim 24, wherein the load balancer is not associated with the virtual server.

29. A method according to claim 24, wherein the load balancer is under control of a system manager of the client.

30. A method according to claim 24, wherein receiving the message comprises receiving a DNS query message.

31. A method according to claim 24, wherein receiving the message comprises receiving from a DNS server.

32. A method according to claim 24, wherein receiving the message comprises receiving a connection establishment request directed to the virtual server.

33. A method according to claim 24, wherein receiving the message comprises receiving a message directed to the load balancer.

34. A method according to claim 24, wherein selecting one of the servers

comprises selecting a server which has a lowest cost path to the load balancer.

35. A method according to claim 24, wherein selecting one of the servers comprises selecting a server which has a lowest delay path or a highest packet size path to the load balancer.

36. A method according to claim 24, wherein the load balancer is geographically closer to the client than to the selected server.

37. A method of selecting a server to be accessed, comprising:

receiving, by a load balancer, a message relating to a virtual server, hosted by a plurality of servers, and to a client desiring to receive data from the virtual server; and

selecting, by the load balancer, one of the plurality of servers to provide data to the client, at least partially responsive to the cost of communications between the client and one or more of the plurality of servers,

wherein the load balancer comprises a client-controlled load balancer that directly selects said one of the plurality of servers representing the virtual server based on said one or more parameters.

38. A method according to claim 37, wherein selecting one of the servers comprises selecting a server under a constraint that a lowest cost client communication connection is used in connecting to the server.

39. A method according to claim 37, wherein selecting one of the servers comprises selecting a server which minimizes a weighted sum of communication costs

to the server and at least one other route related parameter.

40. A method according to claim 39, wherein selecting one of the servers comprises selecting a server which minimizes a weighted sum of the communication costs to the server and the round trip delay to the server.

41. A load balancer, comprising:

an interface adapted to receive server access messages from clients; and

a processor adapted to determine, for at least one of the messages, whether the message requires load balancing responsive to at least one attribute different from the identity of the server referenced by the message, and to select for at least one message determined to require load balancing, a server to service the client,

wherein the processor comprises a client-controlled processor that directly selects the server to service the client based on the at least one attribute.

42. A load balancer according to claim 41, wherein the at least one attribute comprises the time at which the message is received at the interface.

43. A load balancer according to claim 41, wherein the at least one attribute comprises the identity of the client.

44. A load balancer according to claim 41, wherein the at least one attribute comprises a protocol to govern the communication with the server.

45. A load balancer according to claim 41, further comprising a packet changing

unit adapted to change the contents of at least one field of packets belonging to connections for which load balancing was performed.

46. A load balancer according to claim 41, wherein the packet changing unit is adapted to change packets in accordance with half NAT or full NAT procedures.

47. A method of selecting a server to be accessed, comprising:

receiving, by a load balancer, a message relating to a virtual server, hosted by a plurality of servers, and to a client desiring to receive data from the virtual server;

choosing a function from a plurality of predetermined functions utilized by the load balancer for selecting servers, responsive to the received message; and

selecting, by the load balancer, one of the plurality of servers that minimizes or maximizes the chosen function, to provide data to the client,

wherein the load balancer comprises a client-controlled load balancer that directly selects said one of the plurality of servers representing the virtual server that minimizes or maximizes the chosen function.

48. A method according to claim 47, wherein choosing the function comprises choosing responsive to an identity of the client.

49. A method according to claim 47, wherein choosing the function comprises choosing responsive to a time at which the message is received.

50. A method according to claim 47, wherein at least two of the predetermined functions depend on different groups of one or more parameters.

51. A method according to claim 47, wherein at least two of the predetermined functions depend on the same parameters but give different weight to one or more of the parameters on which they depend.

52. A method of selecting a server to be accessed by a client via a wide area network (WAN) from among a plurality of servers associated with a domain name comprising:

providing a client-controlled load balancer in a local area network (LAN) connected to the WAN, the LAN including the client;

receiving at the load balancer a list of addresses of servers hosting the domain name; and

selecting by the load balancer one of the addresses of the plurality of servers based on a parameter related to a path between a point in the vicinity of the client and one of the plurality of servers.

53. The method of claim 52 wherein the parameter is time-variable.

54. The method of claim 52 wherein the parameter comprises a measure of communication quality.

IX. EVIDENCE APPENDIX

(None)

X. RELATED PROCEEDINGS APPENDIX

(None)